J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

The Generalizability Puzzle

Mary Ann Bates
Executive Director, J-PAL North America
Department of Economics, MIT

J-PAL Evaluating Social Programs Webinar Series Agenda

- Day 1: Theory of Change and Measurement: An Interactive Case Study
- Day 2: Why Randomize? An Interactive Case Study
- Day 3: Ethics of Randomized Evaluations
- Day 4: Building Effective Research-Practitioner Partnerships
- **Day 5: The Generalizability Puzzle**

Where participants joined us from on Day 1:

Participant question from yesterday: “Which countries would you be looking to expand this [READI Chicago] programme to?”

"We keep running into the same problem from place to place to place. … The solutions, in a sense, can be the same. **You learn something general**, and from this general finding, you can **extract a lesson that policymakers will then tailor to each individual context**."

Esther Duflo, interview after the announcement of the 2019 Prize in Economic Sciences
https://bit.ly/2WI37Bk

Example:
HIV Relative Risk Information Campaign

A "Relative Risk Information Campaign" in Kenya led to a significant reduction in unwanted teenage childbearing with older partners.

Dupas 2011

Photo: Aude Guerricci, for evaluation "HIV/AIDS Prevention Through Relative Risk Information for Teenage Girls in Kenya"

Randomized evaluation: Relative Risk Information

- Study by Pascaline Dupas (Stanford)
- Location: rural western Kenya
- 71 schools randomly selected from 328 schools
- Trained project staff visited the 8th grade classrooms
  - 10-minute video
  - Detailed stats on the rates of HIV by age and sex from nearby Kisumu
  - 30-minute discussion of cross-generational sex

Men's HIV Rates by Age in Kisumu, Kenya, 2001

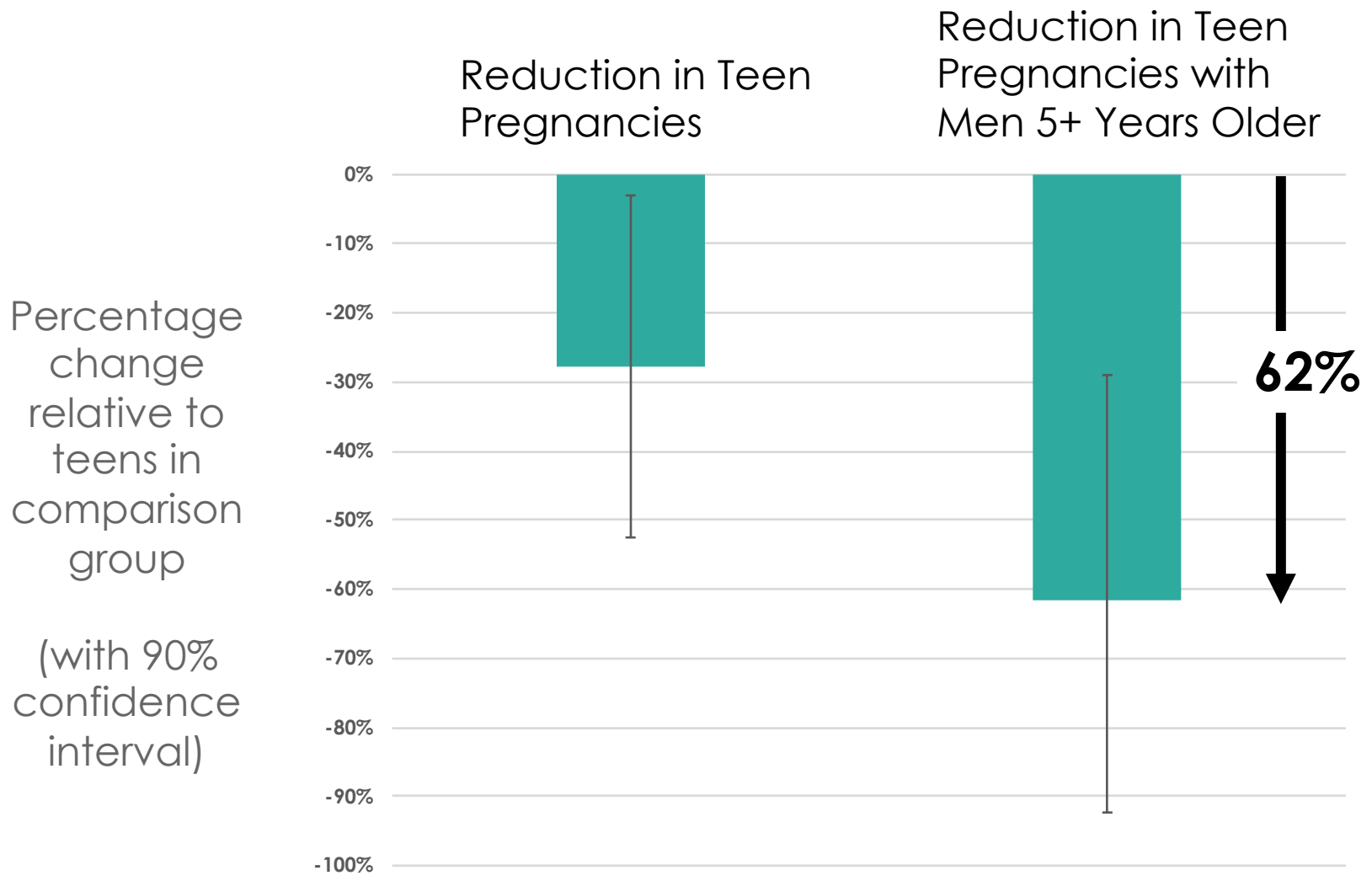
Age	HIV prevalence rate
Age 15-19	4%
Age 20-24	13%
Age 25-29	28%
Age 30-39	32%

Republic of Kenya Ministry of Health, "AIDS in Kenya," 2001.

Example:
HIV Relative Risk Information Campaign

Photo of a relative risk education session in Botswana. young1ove.org

HIV Relative Risk Information Campaign Reduced Teen Pregnancies in Kenya



Should Rwanda replicate the program?

A. Yes

B. No

Should Rwanda replicate the program?

Share some reasons why you said yes or no

The challenge

- Dramatic rise in the number of rigorous impact evaluations in developing and developed countries in last 20 years

- Unlikely to be rigorous evaluation of the program policy makers want to introduce in exactly same location and conditions

Four misguided questions

- Can a study inform policy only in the location in which it was undertaken?
- Should we use only whatever evidence we have from our specific location?
- Should a new local randomized evaluation always precede scale up?
- Must an identical program or policy be replicated a specific number of times before it is scaled up?

- What counts as a "similar enough" new setting?

The generalizability puzzle framework

- Instead of focusing on place and time, focus on people
  - Key conditions and general lessons about behavior
- Evidence from single study just one part of the puzzle
  - We weigh the evidence based on quality and adjust priors
- Combine theory, descriptive evidence, and results of rigorous impact evaluations to answer:
  - Whether results from one country likely to replicate in another
  - When we need more evaluation and when we don't
- For more detail, see Mary Ann Bates and Rachel Glennerster, "The Generalizability Puzzle," *Stanford Social Innovation Review*, 2017. https://ssir.org/articles/entry/the_generalizability_puzzle

Examples of How to Apply the Generalizability Puzzle Framework

Scaling immunization incentives

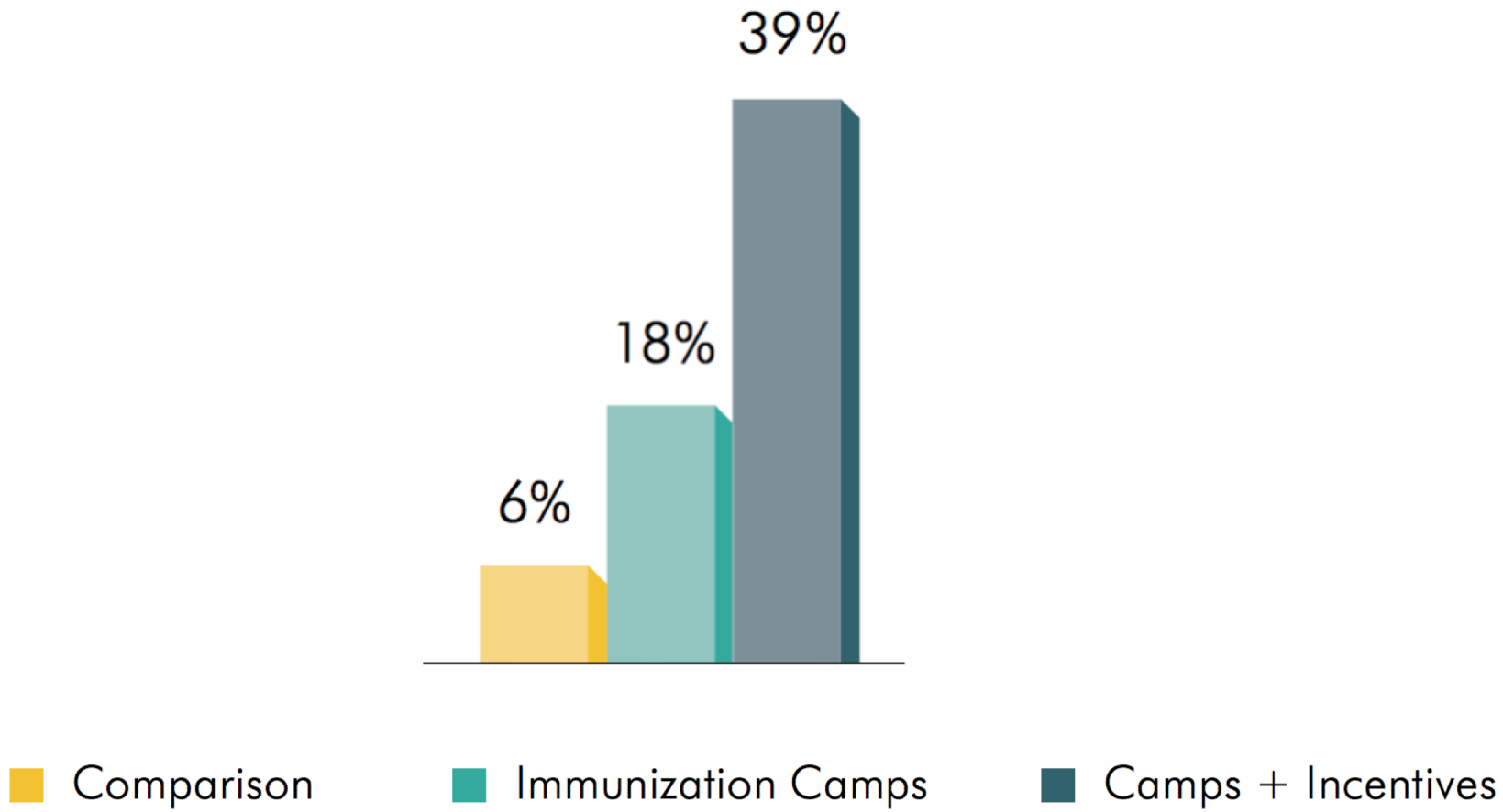
- Seva Mandir program to increase immunization rates in rural Rajasthan, tested with RCT
  - Banerjee, Duflo, Glennerster, Kothari, 2010

- Fixing supply with **reliable infrastructure**: regular monthly immunization camps with nurse present without fail

- Building demand with **incentives**: 1kg lentils for every vaccination, set of plates on completed immunization schedule

A parent receives a kilogram of lentils at a vaccination clinic in Rajasthan, India.

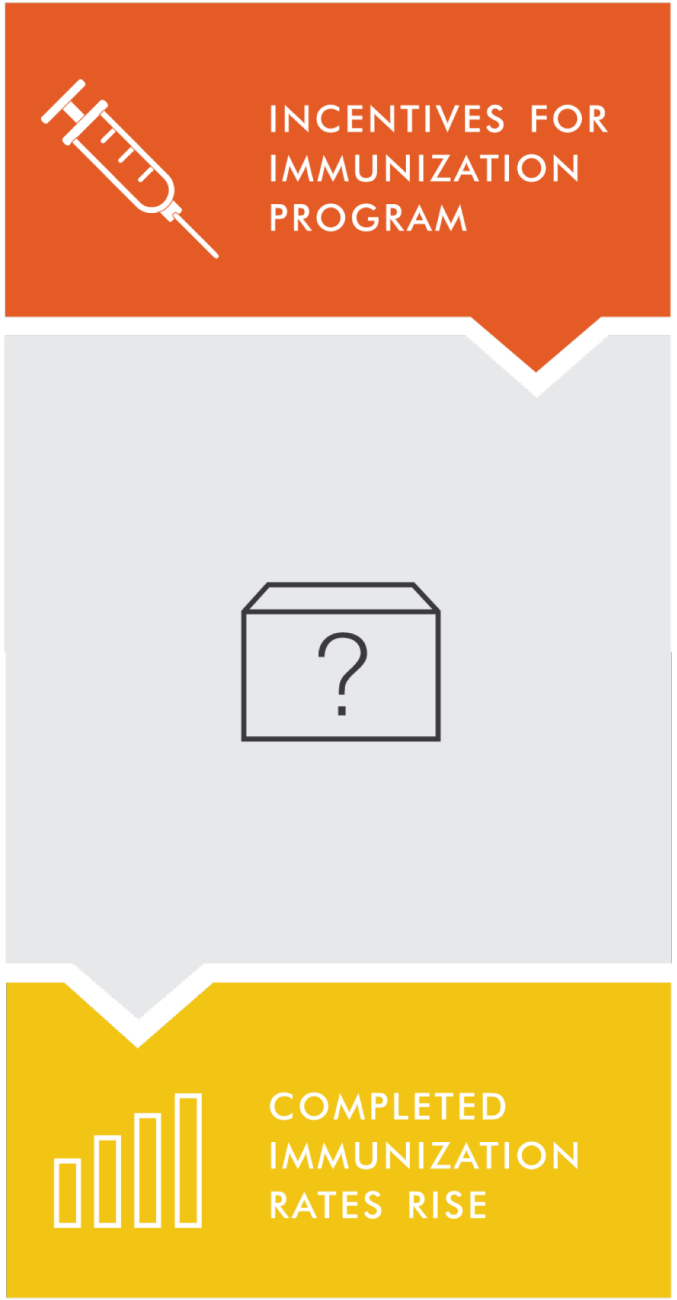
Percentage of Children Aged 1-3 Years Who Have Completed A Course of Immunizations



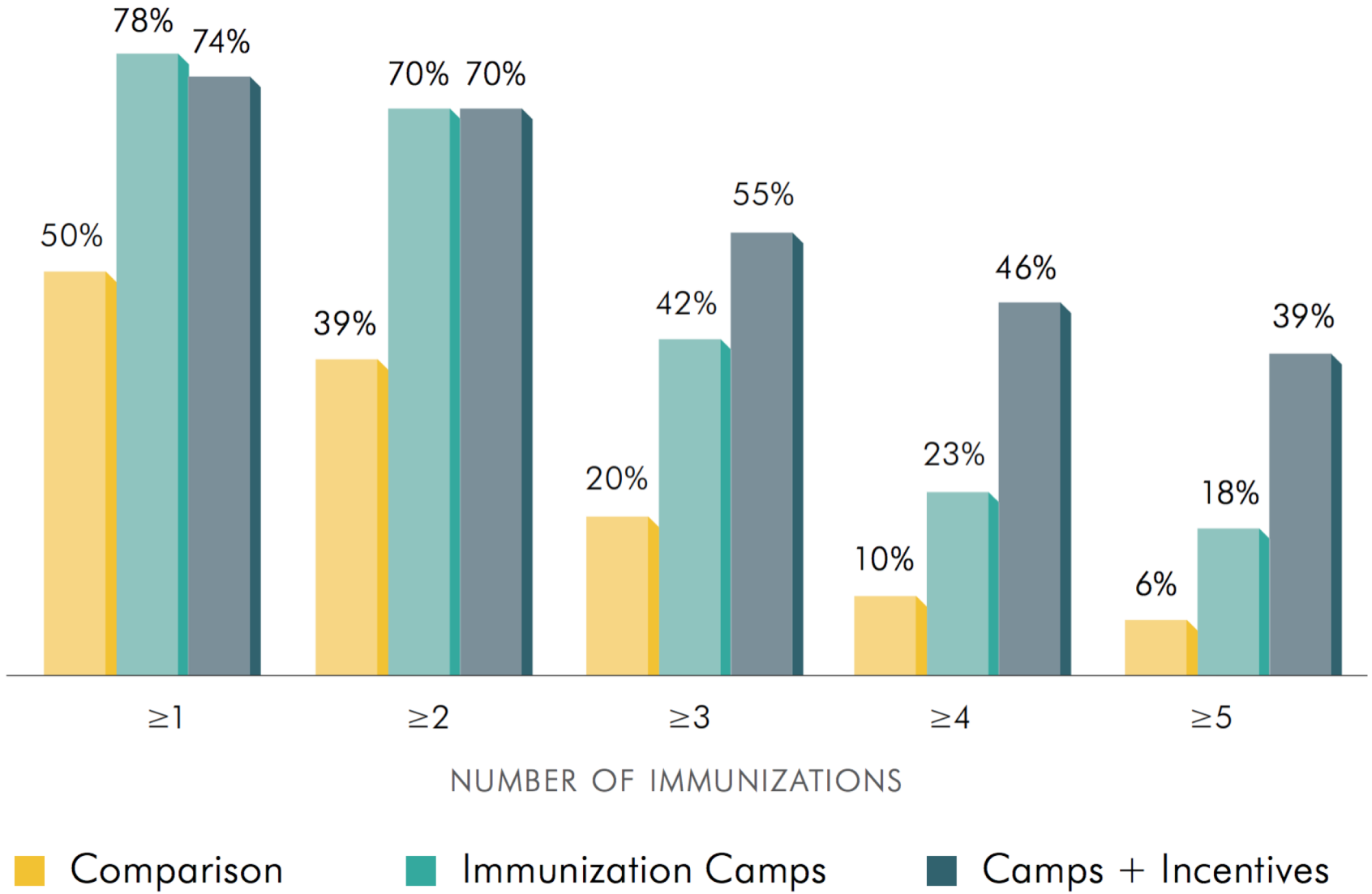
Viewing evidence in isolation

If a government in West Africa wanted to improved immunization rate, should they consider incentives?

- Only one RCT in South Asia; not Africa
- Program conducted by NGO, not government
- Lentils not core part of local diet

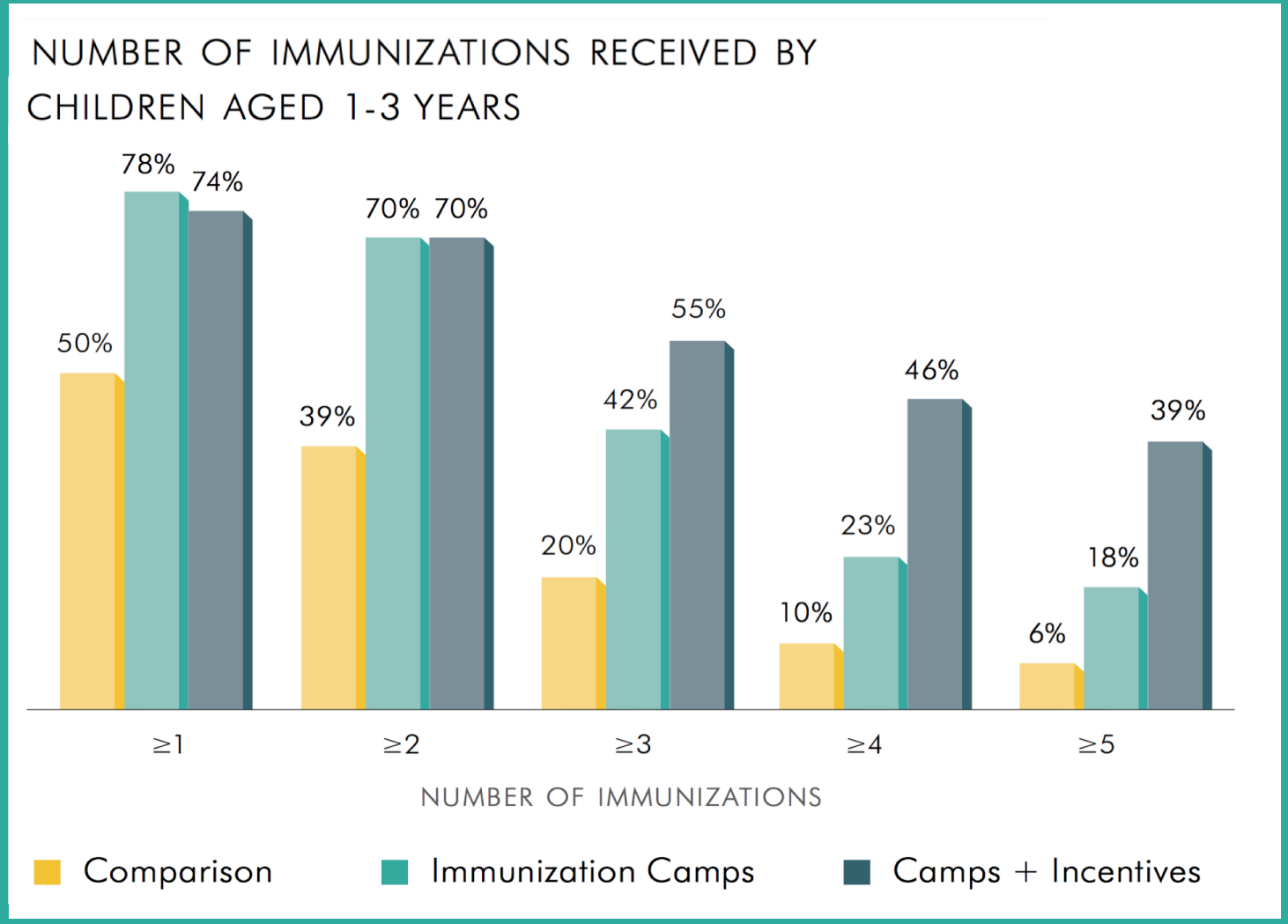


NUMBER OF IMMUNIZATIONS RECEIVED BY CHILDREN AGED 1-3 YEARS



Imagine you are considering replicating or adapting this program

What do you notice about these results?



Generalizability Framework

We all struggle with prevention and procrastination

- People procrastinate and find it hard to stick with behavior they believe is good for them and their children

- Small changes in the prices of preventative products sharply reduce take-up (9+ RCTs)

- Even very small incentives can influence non-trivial decisions
  - Small conditional cash transfers (CCT) can have similar impacts to bigger CCT (Baird et al. 2010, Malawi)
  - Relatively small incentives can be effective at:
  - encouraging HIV testing (Thornton 2008, Malawi)
  - preventing child marriage (Buchmann et al. 2017, Bangladesh)
  - increasing take-up of flu vaccinations (Alsan et al. 2017, United States)
  - combating diabetes (Aggarwal et al. 2020, India)

Generalizability Framework

Which country is a good potential scale-up location for incentives?

A. Country 1

B. Country 2

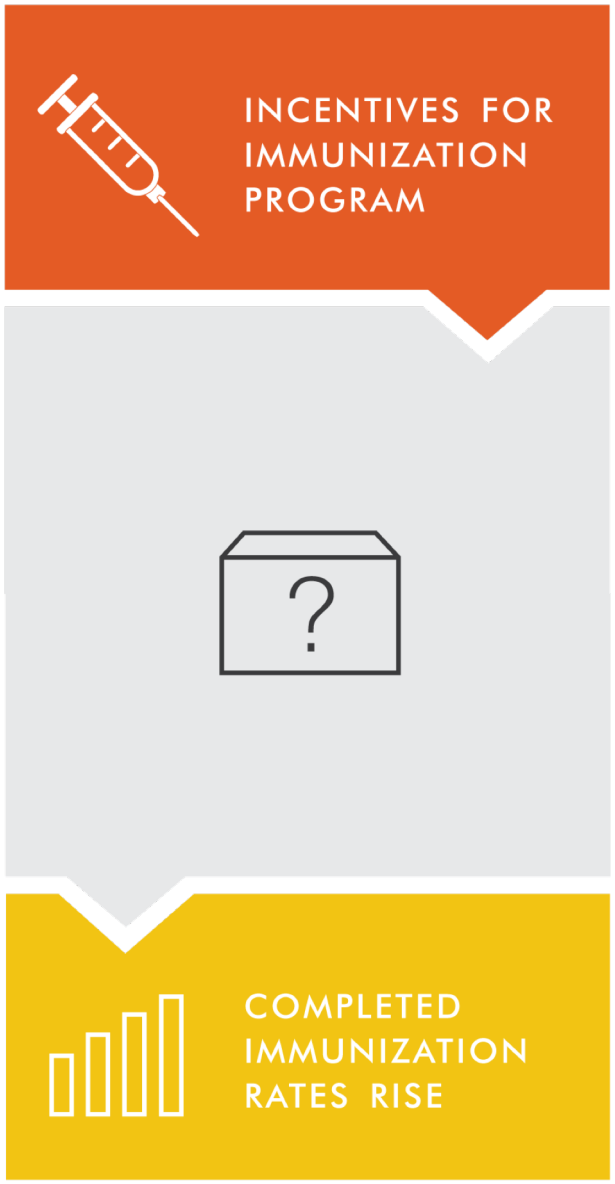
C. Neither

D. Both

Immunization rates by antigen		
	Country 1	**Country 2**
DPT1	84	47
DPT3	74	41
Measles	67	41
Fully immunized	49	38

Local Evidence on Implementation

- This is where the switch from reliable NGO to government delivery will be critical
- Result with a government might be different than with NGO. Should we do an RCT?
- Perhaps test incentives for effective delivery within government


INCENTIVES FOR IMMUNIZATION PROGRAM
?
COMPLETED IMMUNIZATION RATES RISE


INCENTIVES FOR IMMUNIZATION PROGRAM
LOCAL CONDITIONS
1. Parents want to vaccinate
2. Parents can access clinic
3. Provider presence sufficient
4. Full immunization schedule is salient
1. Parents procrastinate or fail to persist
2. Parents are highly sensitive to price of preventative health
GENERALIZED LESSONS ON BEHAVIOR
LOCAL IMPLEMENTATION
1. Incentives delivered to clinics
2. Incentives delivered to parents
COMPLETED IMMUNIZATION RATES RISE

Applying the Generalizability Puzzle Framework

Three examples

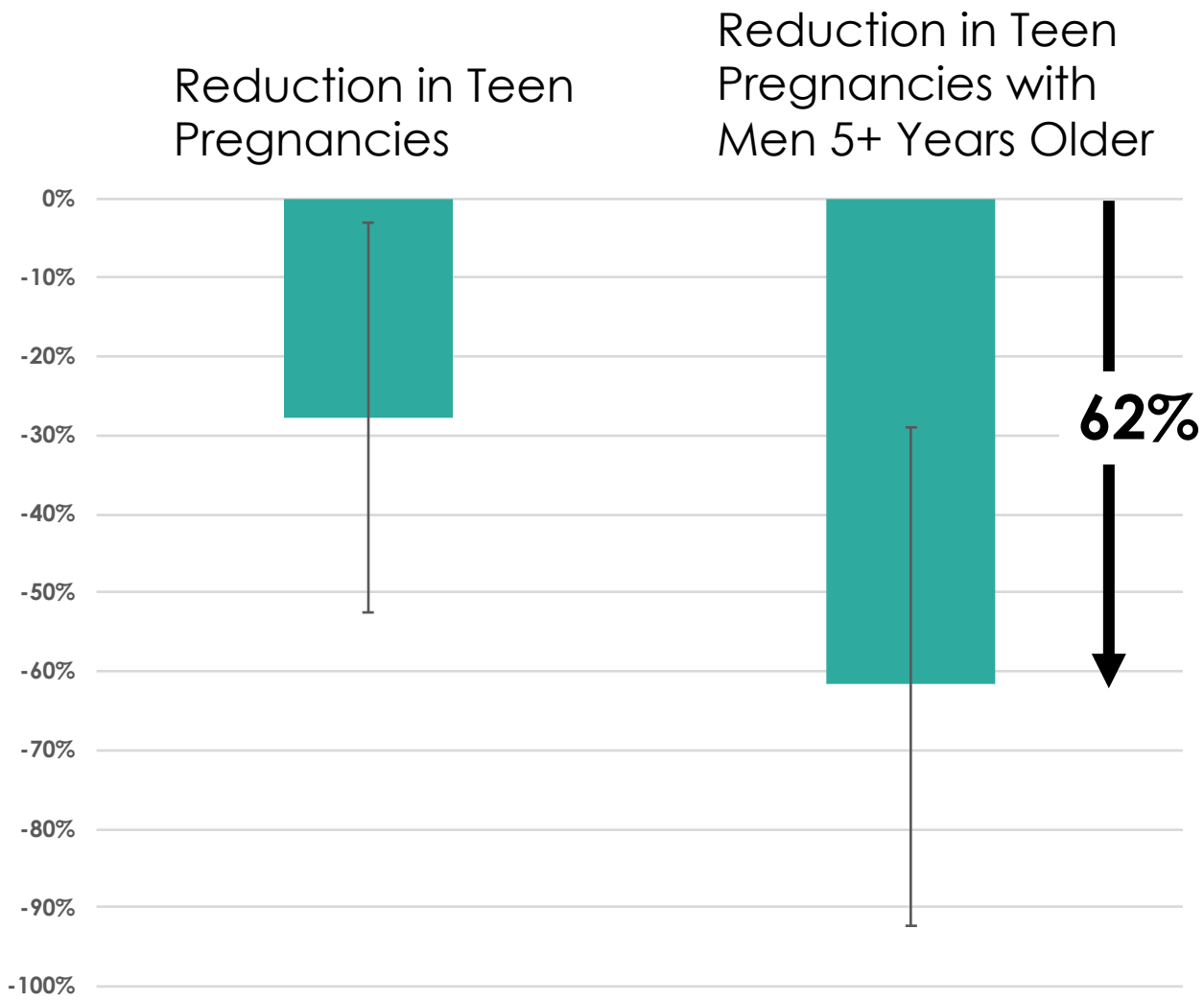
1. Scaling immunization incentives
2. Relative risk education program
3. Teaching at the right level

Would the HIV Relative Risk Information Campaign work in Rwanda?

Results of the study in Kenya:

Percentage change relative to teens in comparison group

(with 90% confidence interval)



Dupas 2011

Generalizability Framework: HIV Relative Risk Program

- What informed teenagers' encounters with sexual partners?
  - Teens knew that unprotected sex can lead to HIV
  - Teens did not know that older men were more likely to be HIV positive than younger men
- Impact of information on behavior depends on how it changes people's prior beliefs
- Key question for scaling is **prior beliefs in the new setting**

What local information would be relevant?

What conditions would need to be similar?

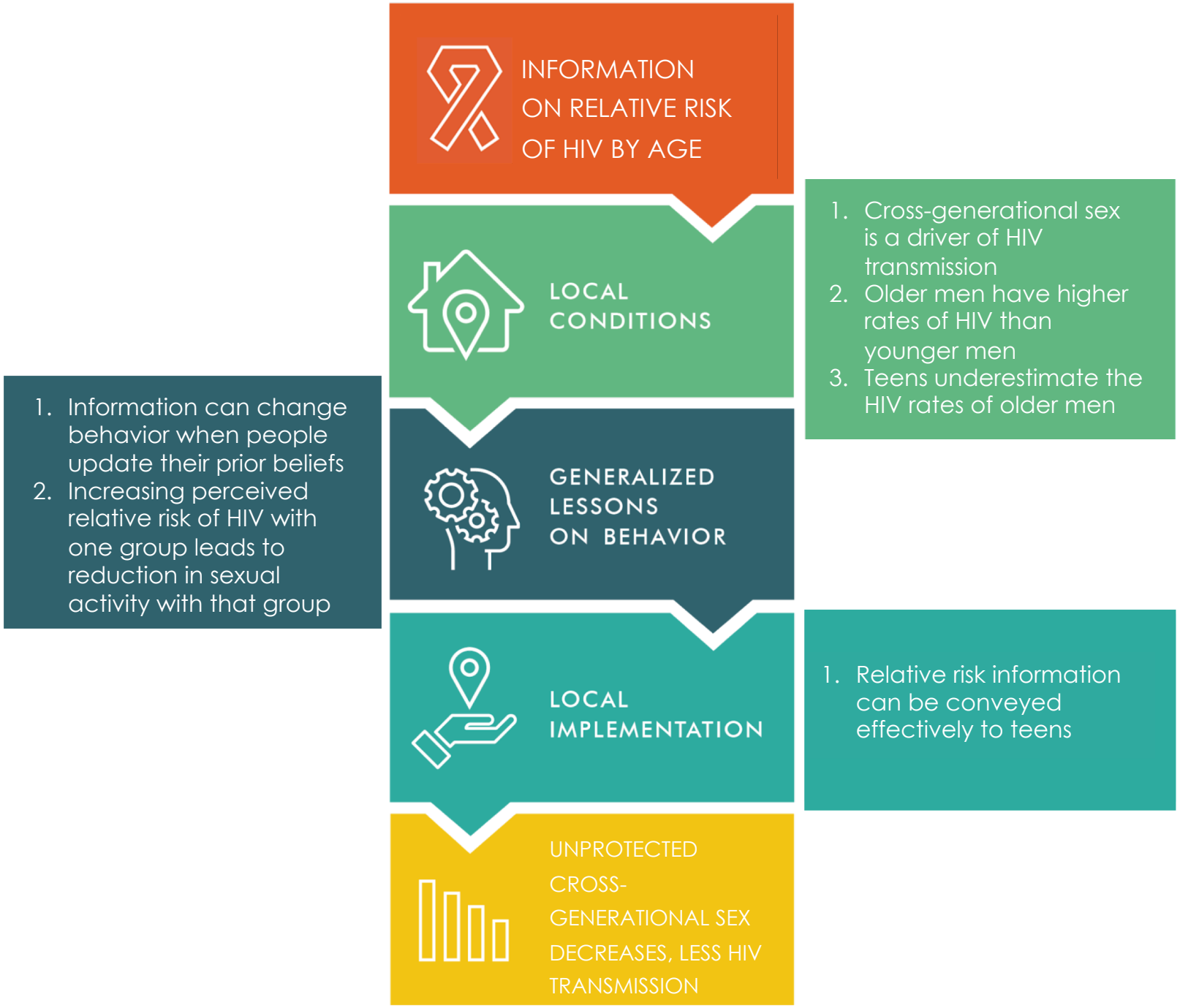
Local descriptive data (collected in a few weeks)

- In Rwanda, men ages 25-29 had an HIV rate of 1.7 percent
- 98% of students overestimated the rate of HIV among men ages 25-29
- **In which direction would a risk awareness program change the Rwandan students' prior beliefs?**

Should Rwanda replicate the program?

A. Yes

B. No


INFORMATION ON RELATIVE RISK OF HIV BY AGE
1. Cross-generational sex is a driver of HIV transmission
2. Older men have higher rates of HIV than younger men
3. Teens underestimate the HIV rates of older men
LOCAL CONDITIONS
1. Information can change behavior when people update their prior beliefs
2. Increasing perceived relative risk of HIV with one group leads to reduction in sexual activity with that group
GENERALIZED LESSONS ON BEHAVIOR
1. Relative risk information can be conveyed effectively to teens
LOCAL IMPLEMENTATION
UNPROTECTED CROSS-GENERATIONAL SEX DECREASES, LESS HIV TRANSMISSION

Young 1ove

EVIDENCE-BASED

We comb academic papers for relevance to our mission, model and niche, and sufficiently rigorous evidence. Our team sifts through jargon, equations and other arcane details tucked away in these papers, and then pulls out and codifies the theory of change behind the proven social impact. We then solicit feedback from experts in the field and put pen to paper, creating evidence-based curriculum. The final step in the translation process is personifying our curricula via trained peer facilitators who deliver our evidence-based messages in partnership with government in schools throughout Eastern and Southern Africa, continuing to learn as we scale.

FOR YOUTH BY YOUTH

We believe proven health and education needs to be taught *by youth for youth*. You can't send old officials to teach kids about sex and stigmatized topics. It doesn't work. It's not relatable. It's boring. It doesn't have impact. We make sure the messages we pick up - the ones research has shown work -- also get delivered in a way our target audience deserves: by youth for youth.

Applying the Generalizability Puzzle Framework

Three examples

1. Scaling immunization incentives
2. Relative risk education program
3. Teaching at the right level

COVID-19 response in education

The School Year Really Ended in March

Abrupt closings have stalled the learning of millions of students. U.S. education needs a rescue, an economist says, and it won't be cheap.

Lily Padula

By Susan Dynarski

https://www.nytimes.com/2020/05/07/business/school-education-online-money.html

If $3x - 10 = 24$, then $x = ?$

For all a and b,
$6a^2b^3 - 3a^2b$ is equivalent to
which of the expressions?

$8 + 14 - 7$

7×4

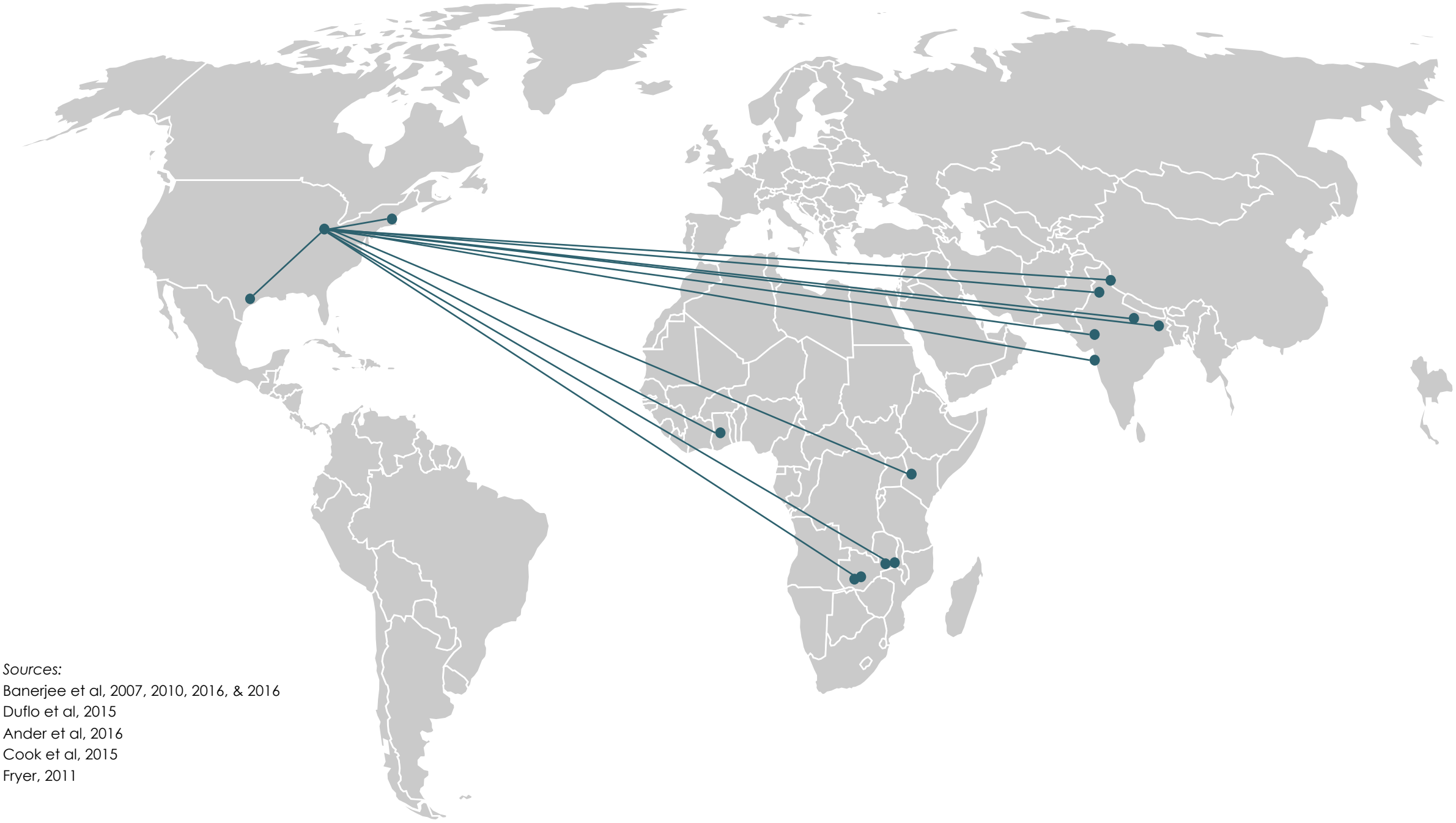
Saga Innovations

J-PAL affiliates and co-authors partnered with Chicago Public Schools to study the impact of Saga Education's model of individualized math tutoring on academic outcomes for 9th and 10th grade male students

- Saga assigned students to a **one-hour tutoring session every day as part of their regular class schedule**.
- **Tutors met with two students at a time** and divided instructional time evenly between **reviewing foundational skills—targeting instruction—and working on current topics from students' regular math classes**.

Results:

- Students in Saga learned an **extra one to two years' worth of math** beyond what their peers learned in an academic year. Tutoring raised participants' average national percentile rank on 9th and 10th grade math exams by more than 20 percent.



Sources:
Banerjee et al, 2007, 2010, 2016, & 2016
Duflo et al, 2015
Ander et al, 2016
Cook et al, 2015
Fryer, 2011

Teaching at the right level

Saga Education tutoring session
www.povertyactionlab.org/case-study/individualized-tutoring-improve-learning

TaRL activities in a classroom in Gujarat, India
www.povertyactionlab.org/case-study/teaching-right-level-improve-learning

Teaching at the right level

The approach works by

- dividing students into groups based on learning needs rather than age or grade;
- dedicating time to basic skills rather than focusing solely on the curriculum; and
- regularly assessing student performance, rather than relying only on end-of-year examinations.

https://www.teachingattherightlevel.org/

Targeted instruction increases learning

Series of studies shows targeted instruction can work in a variety of contexts:

1. Balsakhi Assistant Programme in India (Duflo et al 2007)
2. Read India Programme (Banerjee et al 2007)
3. Computer Assisted Learning (Duflo et al 2007)
4. India Reading Camps (Banerjee et al 2010)
5. Extra Teacher Programme in Kenya (Duflo et al 2011)
6. Haryana Learning Enhancement Programme (Berry et al 2013)
7. TCAI Programme in Ghana (Duflo and Kiessel 2012)
8. Match Education and Youth Guidance in Chicago (Cook et al 2014)
9. Match Education in Boston (Cook et al 2015)
10. Saga Innovations in Chicago (Davis et al 2017)

For more, see: J-PAL Evidence Review. 2019. "Will Technology Transform Education for the Better?"

TEACHING AT THE RIGHT LEVEL

Zambia

This case study shares J-PAL Africa and Pratham's experience of working with the Ministry of General Education in Zambia to build the Catch Up programme, where global evidence, local adaptation, and iterative testing were used to improve learning outcomes.

TARGETED INSTRUCTION / TUTORING PROGRAM

LOCAL CONDITIONS

1. Children attend school, but literacy and numeracy rates are low
2. Teachers face incentives to teach grade-level material, not catch-up material

1. Catch-up program instruction is at the student's level
2. Students learn when material is at their level

GENERALIZED LESSONS ON BEHAVIOR

LOCAL IMPLEMENTATION

1. Teachers/volunteers trained in catch-up program
2. Time is devoted to catch-up program
3. Students attend catch-up classes targeted to their learning level

LITERACY AND NUMERACY RATES RISE

Imagine that **you are the superintendent of a large school district**, and are looking for ways to boost student performance during and after periods of remote learning.

You recently heard about Saga's program for teaching at the right level, and want to explore whether it makes sense for you to implement this program in your schools.

Is evidence on teaching at the right level relevant to your COVID-19 response?

How might you apply this evidence?

What data would you use to find out if students are performing below grade level and if there are varying levels of achievement?

What data would you use to find out whether teachers may be teaching at one level for all the students in their classroom?

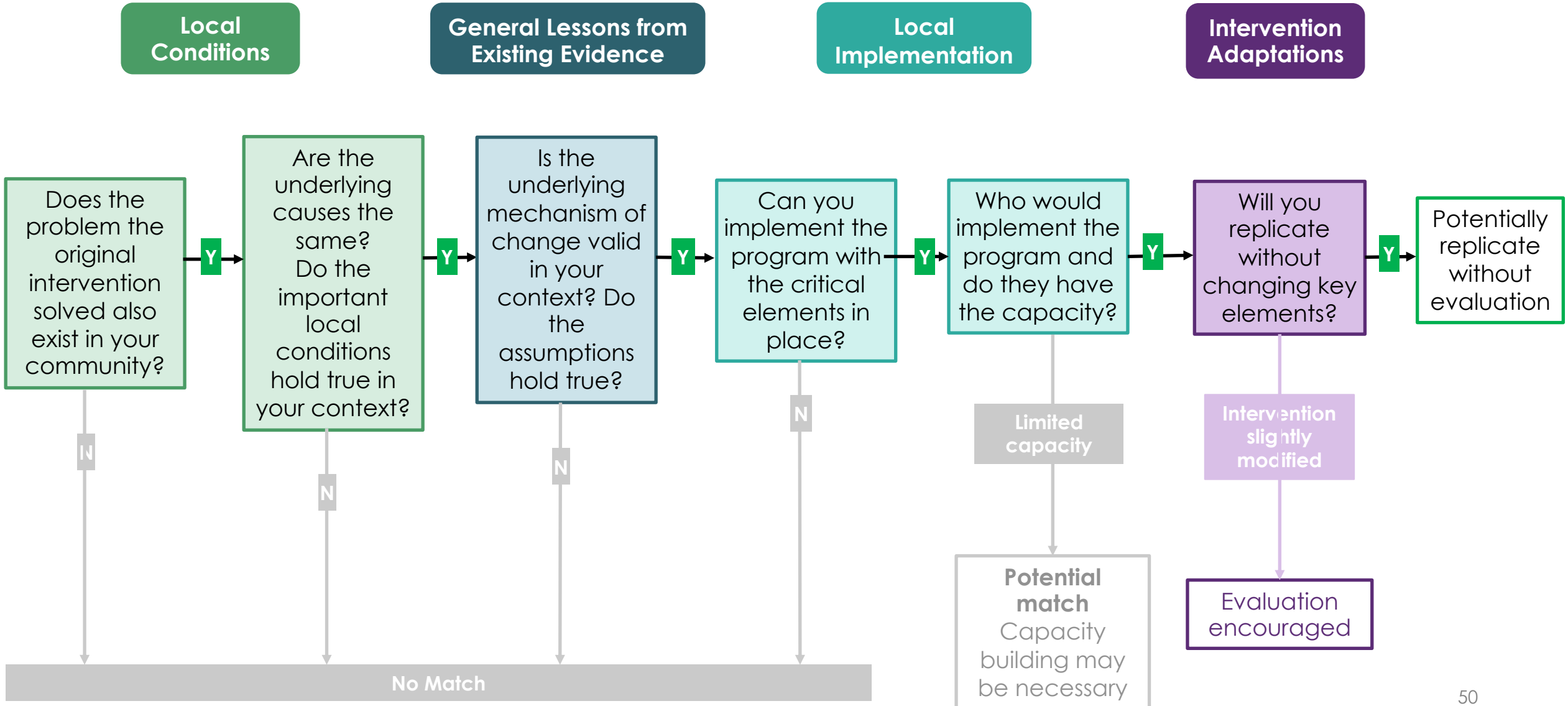
What data would you use to decide which students to focus on with a Teaching at the Right Level program?

Indicators and Data for Decision-Making

What metrics and data would you use to assess whether the important local conditions hold in your school district? How would you determine what grades and students to target?

Local Conditions	Indicators	Data Sources
1. Students (at least some) are **performing below grade level**	- *Standardized test scores* - *Teachers' evaluations of student achievement*	- *School and state databases of standardized tests* - *Interviews with teachers*
2. There are **varying levels of student achievement** in classrooms, with some students performing above, at, and below grade level.	- *Grades* - *Standardized test scores* - *Teachers' evaluations of student achievement*	- *School grading systems* - *State standardized tests* - *Interviews with teachers*
3. Teachers **teach at one level for all students** in their classroom, for either practical reasons or the school's incentive structure.	- *Teachers' assessments of their own teaching* - *Schools' incentives structures for teachers* - *Curriculum*	- *Interviews with teachers* - *Records about schools' incentives structures for teachers* - *Records on curriculum*

A Decision Tree



Conclusion

Does evidence from RCTs replicate to new contexts?
Too big a question. Break it down:

- What is the theory of change behind the RCT?
- Do the local conditions hold for that theory to apply?
- How strong is the evidence for the general behavioral change?
- What is the evidence that the implementation process can be carried out well?

Conclusion

- If we have enough evidence to act, do we have enough evidence to stop evaluating impact? (Always monitor)
  - We often need to act even when evidence is thin
- Often big overlap between when have enough evidence to launch a new initiative and when it is still worth evaluating
  - Questions may remain about best way to implement
- Tradeoff between evidence in new areas, versus more on improving evidence on refining a program

Are the locations identical?

Is there a similar problem?

Why did a solution work?

Icon by Adrien Coquet from the Noun Project

Over 400 million people reached by scaling up programs found to be effective by J-PAL RCTs

Evidence to Policy

Evidence to Policy

Evidence from randomized evaluations is changing how we understand and address problems related to poverty. Policymakers, practitioners, and funders worldwide are increasingly applying this learning to social policies and programs.

Over 400 million people have been reached by programs that were scaled up after being evaluated by J-PAL affiliated researchers. Many more have benefitted from the several broader ways evidence can inform policy, outlined below.

Continue reading

Pathways to Policy Change

Below, you will find six pathways through which evidence can have an impact on policy and case studies that illustrate partnerships leading to policy impact.

Shifting global thinking

Knowledge generated by randomized evaluations has fundamentally shaped our understanding of many social policies.

Example case studies:

Free bednets to fight malaria

More...

Institutionalizing evidence use

Many organizations, including governments and large NGOs, have institutionalized processes for rigorously evaluating innovations and incorporating evidence into decision-making.

Example case studies:

A government innovation lab to improve education

More...

Applying research insights

Lessons from randomized evaluations have informed the design of programs.

Adapting and scaling a program

Programs originally evaluated in one context have been adapted and scaled in others.

J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

Further reading and resources

- Bates and Glennerster, 2017, "The Generalizability Puzzle," *Stanford Social Innovation Review* https://ssir.org/articles/entry/the_generalizability_puzzle
- Kremer and Glennerster, 2012, Chapter in *Handbook of Health Economics*
- J-PAL Evidence to Policy page http://www.povertyactionlab.org/evidence-to-policy/
- J-PAL Self-Guided Case Study on Applying the Generalizability Framework to Complex Health Care https://www.nationalcomplex.care/research-policy/resources/toolkits/case-study-generalizability-framework/